counts, to question the interpretation which attributes red-shift to recession.

The result can best be shown numerically, through least-squares solutions for m_1 and B, which are related (when we adopt the velocity-distance factor derived at Mount Wilson) by the formula[8]

$$m_1 = m_0 - 1.667 \log \overline{N}_m - 10^{0.2\,(m_0 - \Delta m_0) + \log B - 4.707}$$

where m_0 and Δm_0 are the observed magnitude limit and the red-shift effect. We use all of the survey material presented by Hubble (Section 6 above) except that we substitute the Harvard counts to $18^m.2$ (nine inner squares only) for his data at the brighter limit, and obtain the following results for the eastern and western parts of the galactic cap:

	EAST	WEST
$\log \overline{N}_m$	1.784	1.635
m_1	15.01	15.26
B	3.48	1.81

If we should assume that the limit of completeness of the Harvard survey is 18.1 instead of 18.2, we find:

	EAST	WEST
m_1	14.90	15.19
B	4.2	2.2

[1] See, for example, these PROCEEDINGS, **22**, 621–627 (1936).

[2] Detailed tabulations are given in Harvard Circular 423 (*in press*).

[3] *Harv. Ann.*, **105**, No. 8, p. 137 (1937).

[4] *Lick Obs. Bull.*, **16**, 177 (1934).

[5] *Mount Wilson Contr.* No. 557, p. 15 (1936). Hubble's C is $-0.6\, m_1$.

[6] *Mon. Not. R. A. S.*, **97**, 156 (1937).

[7] *Ibid.*, **97**, 506 (1937).

[8] The form of the exponential term is suggested by Hubble.

DEPENDENT PROBABILITIES AND SPACES (L)

BY GARRETT BIRKHOFF

HARVARD UNIVERSITY

Communicated February 1, 1938

1. *Introduction.*—The purpose of the present paper is to express the theory of dependent probabilities in new terms.

These terms are: (*a*) the general theory of linear operators on Banach space, and (*b*) the relation $f < g$. The abstract theory of partially ordered Banach spaces has already been formulated by Kantorovitch;[1] it involves such notions as: upper bound, lower bound, least upper bound or sup, greatest lower bound or inf, lim sup, lim inf, positive part, negative part, absolute value and disjointness.

Firstly, the fundamental definitions are stated in a form which includes all known cases; this has never been done before. Secondly, Markoff's fundamental theorem on "probabilities in chain" is proved in a form including all known cases. And lastly, a new theorem, which specializes in the *deterministic* case to von Neumann's well-known Mean Ergodic Theorem,[2] is proved with added generality in the *stochastic case*.

The present paper merely sketches the proofs, which will be given in full elsewhere.

2. *Postulates.*—The model with which we shall work is described in the following:

DEFINITION 1: *By a space* (L), *we mean any space* Σ *which satisfies the following six postulates.*

P0: Σ is a linear space with real scalars, and a relation $f > 0$ (to be read, f is positive), is defined on Σ.

P1: If $f > 0$ and $g > 0$, then $f + g > 0$.

P2: If $f > 0$ and λ is a scalar, then $\lambda > 0$ implies $\lambda f > 0$ and conversely.

P3: Relative to the definition, $f > g$ (read, f is greater than g) means $f - g > 0$, Σ is a *lattice*.[3]

P4: A "norm" $\eta(f)$ is defined on Σ, relative to which Σ is a Banach space.[4]

P5: Norm is additive on positive elements; $f > 0$ and $g > 0$ imply $\eta(f+g) = \eta(f) + \eta(g)$.

Example: Let B denote the Boolean algebra of all subsets X of a class I_n of n points, or of Borel subsets of an interval I (or of any region!) modulo sets of measure zero. Then the additive, continuous functions defined on B satisfy (P0)–(P5), if by $f > 0$ we mean $f(X) \geqq 0$ for all X and $f(I) > 0$, and by $\eta(f)$ we mean $f(I)$ when $f > 0$ and sup $|f(X)| + |f(X')|$ in general Thus the space (L) and its finite-dimensional analogues are "spaces (L)."

Consequences of (P0)–(P2): Define $f > g$ to mean $f - g > 0$. Then (1) Σ is a partially ordered set in the sense of Hausdorff, (2) translations $x \rightarrow x + a$ preserve order, (3) homothetic expansions $x \rightarrow \lambda x$ preserve or invert order according as $\lambda > 0$ or $\lambda < 0$. We note that by (1)–(2), (P3) follows from (P0)–(P2) and the assumption that any element f has a "positive part" $f^+ = f \smile 0$, such that $x \geqq f^+$ implies $x \geqq 0$ and $x \geqq f$ and conversely.

Consequences of (P0)–(P3): If we set $f^+ = f \smile 0$, $f^- = f \frown 0$, and $|f| = f^+ - f^-$, then (4) the "Jordan decomposition" $f = f^+ + f^-$ holds, whence $f + g = (f \smile g) + (f \frown g)$ by (2), (5) $f^+ \frown (-f^-) = 0$—in words, f^+ and $-f^-$ are *disjoint*, (6) the dual distributive laws $f \smile (g \frown h) = (f \smile g) \frown (f \smile h)$ and $f \frown (g \smile h) = (f \frown g) \smile (f \frown h)$ are valid, (7) the triangle law on absolute values holds: $|f - g| + |g - h| \geqq |f - h|$, (8) the functions $f \smile g$ and $f \frown g$ are *monotone* in both variables, and *uniformly continuous* in that $|(f \smile g) - (f^* \smile g)| \leqq |f - f^*|$ and $|(f \frown g) - (f^* \frown g)| \leqq |f - f^*|$.

Consequences of (P0)–(P5): (9) The functions $f \cup g$ and $f \cap g$ are metrically continuous, in virtue of (8), (10) the functional $\lambda(f) = \eta(f^+) - \eta(f^-)$ is *linear*, and $\eta(f) = \lambda(|f|)$, (11) every set of elements of Σ having an upper bound has a least upper bound (and dually).

These results can be found in Kantorovitch, op. cit.

3. *Normal Subspaces and Decompositions.*—In proving our Mean Ergodic Theorem (Theorem 3), but not elsewhere, we shall want two further definitions, which seem to be new.

DEFINITION 2: *A subspace of a linear space satisfying* (P0)–(P3) *is called "normal" if and only if it contains* (a) *with any* f, *also* $|\mathrm{f}|$, *and* (b) *with any positive* f, *all "parts"* x *of* f (*i.e., all positive* x *with* $\mathrm{x} < \mathrm{f}$).

DEFINITION 3: *By a "direct decomposition" of* Σ, *is meant a choice of complementary normal subspaces—that is, of subspaces* S *and* T *such that* $\mathrm{S} \cap \mathrm{T} = 0, \mathrm{S} + \mathrm{T} = \Sigma$.

Remarks: (1) The normal subspaces of Σ correspond to its homomorphisms in just the same way that the normal subgroups of a group correspond to *its* homomorphisms, and (2) the decompositions of Σ correspond one–one to its representations as a direct union.

4. *Connections with Dependent Probabilities.*—Spaces (L) are connected with the theory of dependent probabilities by three fundamental definitions.

DEFINITION 4: *By a "distribution" is meant a positive element of* Σ *with norm one.*

DEFINITION 5: *By a "transition operator" on* Σ *is meant an additive operator which carries distributions into distributions.*

DEFINITION 6: *A transition operator* T *describing the dependence of the state of a system at time* t′ *on its state at a previous instant* t, *is called "independent" of an operator* U *relating the instants* t″ *and* t′ [$\mathrm{t}'' > \mathrm{t}'$], *if and only if the instants* t″ *and* t *are related by the transition operator* TU: $\mathrm{f} \to (\mathrm{fT})\mathrm{U}$.

As authorities for these definitions, we can cite the usual formulations of Bayes' Theorem, of the theory of Markoff chains, of Kolmogoroff's more general theory of "stochastic processes." Also, Fourier's theory of heat flow is expressed by transition operators: the invariance of $\lambda(f)$ is the gist of the first, and that of the set of $p \geqq 0$ of the second, law of thermodynamics. Finally, the flows of phase-space envisaged by Poincaré in his version of classical mechanics, induce automorphisms on the space (L)—and hence are transition operators in our sense, as well as "unitary operators on Hilbert space."

Conclusions: (1) The set Δ of all distribution functions is a closed convex subset of Σ, of diameter sup $|(p - q)| \leqq 2$, (2) the distance $\eta(|p - q|)$ is the "stochastic distance" recently defined by Mazurkiewicz—it is *not* equivalent to the traditional notion of "convergence in probability," (3)

$|fT| \leqq |f|$, whence $\eta(fT) \leqq \eta(f)$ (T is of "modulus" unity, and a "contraction," and so uniformly continuous!).

5. *Hypothesis of Markoff.*—Now let Σ be any space (L), and T a fixed transition operator on Σ.

DEFINITION 7: *An* f *in* Σ *is called a "fixpoint" if and only if* $fT = f$. *A distribution which is a fixpoint is called "stable."*

Results: (1) The fixpoints are a closed linear subspace of Σ. Hence the stable distributions are a closed convex subset of Δ, and their number is either zero, or one (the "metrically transitive" case[5]), or infinity. All three cases are possible, but the second is the most interesting.

Hypothesis of Markoff (weakened): For some n, $d = \inf_{p\epsilon\Delta} pT^n > 0$.

THEOREM 1: *If* T *satisfies Markoff's hypothesis, then there is a unique stable distribution* p_0. *Moreover the* pT^k *tend to* p_0 *uniformly, with the rapidity that the terms of a convergent geometrical progression tend to zero.*

Proof: First, $\eta(pT^n - qT^n) \leqq (1 - |d|)\eta(p - q)$ for any $p, q \epsilon \Delta$. The conclusion now follows by a generalization to complete metric spaces (like Δ) of a simple argument due to Carl Neumann and often exploited by Picard, which is purely geometrical.

COROLLARY 1: $\operatorname{Sup} pT^n \leqq p + \sum_{k=0}^{\infty} (pT^{k+1} - pT^k)$ *is finite, for any fixed* p. (Cf. §2, conclusion (11)).

COROLLARY 2: *Let* $T_1, \ldots, T_n$ *be any sequence of transition operators, and let* d_i *denote* $\operatorname{Inf}_{p\epsilon\Delta} pT_i$. *Then for all* p, q *in* Δ, $\eta(pT_1 \ldots T_n - qT_1 \ldots T_n) \leqq 2\prod_{i=1}^{n}(1 - |d_i|)$.

6. *Ergodic Hypothesis.*—Unless the conclusion of Corollary 1 holds, the means of the pT^k at best tend to 0. Hence we shall fix p, and make the

ERGODIC HYPOTHESIS: The pT^k have an upper bound. This is fulfilled if p is the integral of a bounded density-function, and T leaves measure invariant: the integral of the upper bound to the density function is an upper bound to the pT^k.

THEOREM 2: *The Ergodic Hypothesis implies the existence of at least one stable distribution.*

Proof: Form $h = \operatorname{Lim} \sup_{k\to\infty} pT^k = \operatorname{Inf}_n \sup_{k\geqq n} pT^k$; evidently $hT = T$, $|h| \geqq 1$, and so $h > 0$. Hence $h/|h|$ is a stable distribution.

THEOREM 3 (mean ergodic theorem): *The Ergodic Hypothesis implies that the means* $\frac{1}{n}\sum_{k=0}^{n-1} pT^k$ *converge* weakly, *in the sense that if* $\lambda(f)$ *is any linear functional, then the numerical means* $\phi_n(p) = \lambda\left(\frac{1}{n}\sum_{k=0}^{n-1} pT^k\right)$ *converge in the ordinary sense.*

Proof: By a generalization of a Lemma of Hahn (cf. §7), λ is the sum of its positive and negative parts. Again, by a simple Lemma of Banach (op. cit., p. 54) each part is a constant multiple of a functional satisfying $0 \leqq \lambda(f) \leqq \eta(f)$ for all $f > 0$. Hence we need only consider this case. Again, if k is large, then $h = \lim \sup_{k \to \infty} pT^k$ contains an arbitrarily large part of pT^k, together with all transforms of this part; hence we can assume $p \leqq h$, where $hT = h$.

We shall make these assumptions, and in the proof, shall treat all "parts" f of h on the same footing. First, define $\bar{\phi}(f) = \lim \sup_{n \to \infty} \phi_n(f)$, $\underline{\phi}(f) = \lim \inf_{n \to \infty} \phi_n(f)$. Clearly $0 \leqq \underline{\phi}(f) \leqq \bar{\phi}(f) \leqq \eta(f)$; clearly also $\underline{\phi}(fT) = \underline{\phi}(f)$ and $\bar{\phi}(fT) = \bar{\phi}(f)$. The functionals $\underline{\phi}$ and $\bar{\phi}$ are monotone; they need *not* be linear, but $\underline{\phi}$ is convex while $\bar{\phi}$ is concave. Hence the functionals (we use a construction of F. Riesz)

$$(\bar{\alpha}f) = \sup \sum_i \bar{\phi}(f_i) \qquad \underline{\alpha}(f) = \inf \sum_i \underline{\phi}(f_i)$$

where the summations are with respect to all decompositions of f into (finite or countable; it makes no difference) parts f_i, are, respectively, the least linear functional $\geqq \bar{\phi}$, and the greatest linear functional $\leqq \underline{\phi}$. Moreover $\bar{\alpha}(fT) = \bar{\alpha}(f)$ and $\underline{\alpha}(fT) = \underline{\alpha}(f)$; the functionals are invariant.

Now since $0 \leqq \underline{\alpha} \leqq \underline{\phi} \leqq \bar{\phi} \leqq \bar{\alpha} \leqq \eta$, and $h \geqq f$, in order to conclude $\underline{\phi}(f) = \bar{\phi}(f) = \mathrm{Lim}_{n \to \infty} \phi(f)$, we need only show that $\underline{\alpha}(f) = \bar{\alpha}(f)$, whence, since $\bar{\alpha} - \underline{\alpha}$ is non-negative and linear, we need only show $\bar{\alpha}(h) \leqq \lambda(h) \leqq \underline{\alpha}(h)$. By duality, we need only show $\bar{\alpha}(h) \leqq \lambda(h)$. This is just what we shall prove.

7. *Extension of a Lemma of Hahn.*—For it, we shall need an extension of a Lemma of Hahn.[6] Let $\lambda(x)$ be any linear functional on Σ, let Λ^+ denote the set of $u > 0$ such that $0 < x \leqq u$ implies $\lambda(x) > 0$, and define Λ^- dually. Further, let Λ^0 denote the set of $u > 0$ such that $0 < x \leqq u$ implies $\lambda(x) = 0$.

LEMMA: *Σ is decomposed into three components: a component Λ^+ on which $x > 0$ implies $\lambda(x) > 0$, a component Λ^0 on which $\lambda(x) = 0$, and a component Λ^- on which $x > 0$ implies $\lambda(x) < 0$*

COROLLARY 1: *Any linear functional can be resolved into its positive part and its negative part.*

8. *Completion of Proof.*—Choose $\epsilon > 0$, and denote by g_n the component of h (cf. §7) on which $\phi_n(x) - \bar{\alpha}(x) + \epsilon\eta(x)$ is non-negative. Then irrespective of ϵ,

LEMMA: *The join u of the g_n is h.*

Proof: Consider $r = h - u$; evidently for all n, $r \leqq h - g_n$. Hence $\bar{\phi}(x) \leqq \sup \phi_n(x) \leqq \bar{\alpha}(x) - \epsilon\eta(x)$ for all $x \leqq r$, and so $\bar{\alpha}(r) = \sup \Sigma\bar{\phi}(x_i) \leqq \bar{\alpha}(r) - \epsilon\eta(r)$, whence $\eta(r) = 0$ and $r = 0$.

Now let h_n denote the part of g_n not in $g_1 \cup \ldots \cup g_{n-1}$; the h_n are the components of a direct decomposition of h. Choose M so large that if h^* denotes $h - \sum_{k=1}^{m} h_k$, then $|h^*| < \epsilon$. We shall show that for all $N > M$,

$$\text{(E)} \quad N\lambda(h) \geqq N\bar{\alpha}(h) - N\epsilon\eta(h) - N\epsilon - M\bar{\alpha}(h)$$

from which, dividing through by N, and letting $N \to \infty$, we will get $\lambda(h) \geqq \overline{\alpha}(h) - \epsilon\eta(h) - \epsilon$. Now letting $\epsilon \to 0$, the proof is complete.

The proof of (E) reproduces the combinatorial essence of G. D. Birkhoff's proof of the Strong Ergodic Theorem.[2] We rely on the fact that the "components" of h form a Boolean algebra, and may be treated like sets.

[1] L. Kantorovitch, "Lineare halbgeordnete Raume," *Math. Sbornik*, **2**, 121–68 (1937). Cf. also H. Freudenthal, "Teilweise geordnete Moduln," *Proc. Akad. Wet. Amsterdam*, **39**, 641–51 (1936).

[2] J. von Neumann, "Proof of the Quasi-Ergodic Hypothesis," these PROCEEDINGS, **18**, 70–82 (1932). Our method is that used by G. D. Birkhoff for his stronger result; cf. "Proof of a Recurrence Theorem for Strongly Transitive Systems, and Proof of the Ergodic Theorem," these PROCEEDINGS, **17**, 650–60 (1931).

[3] In the sense of the author's "On the combination of subalgebras," *Proc. Camb. Phil. Soc.*, **29**, 441–64 (1933). Synonyms are "Verband" (Fr. Klein) and "structure" (O. Ore). We shall use the notation $f \cup g$ for sup (f, g) and $f \cap g$ for inf (f, g).

[4] S. Banach, "Théorie des opérations linéaires," Warsaw, 1933. By general consent, the "B-spaces" of Banach, op. cit., are called Banach spaces; they are complete, metric, linear spaces.

[5] In the sense of G. D. Birkhoff and Paul Smith, "Structure Analysis of Surface Transformation," *Jour. Math.*, **7**, 365 (1928).

[6] The author is much indebted to J. von Neumann for suggesting that this lemma could be generalized. He is also indebted to S. Ulam for many conversations on the whole subject.

ON SEMI-GROUPS OF TRANSFORMATIONS IN HILBERT SPACE

BY EINAR HILLE

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated February 9, 1938

1. Let E be a normed complete linear vector space, in other words a space (B) in the terminology of Banach. Let $x \epsilon E$ and let $T_\alpha(x)$ be a linear transformation on E to E defined for every $\alpha > 0$. If

$$T_\alpha(T_\beta(x)) = T_{\alpha+\beta}(x), \tag{1}$$

we say that $\{T_\alpha(x)\}$ forms a *semi-group*.

Assuming in addition that

$$\| T_\alpha(x) \| \leqq \| x \|, \tag{2}$$

I have investigated some of the properties of these transformations.[1] Continuing this study, I have found the case in which E is a Hilbert space and $T_\alpha(x)$ is a self-adjoint, positive definite transformation of particular